REMARKS

The above amendments to the above-captioned application along with the following remarks are being submitted as a full and complete response to the Official Action dated September 7, 2004. In view of the above amendments and the following remarks, the Examiner is respectfully requested to give due consideration to this application, to indicate the allowability of the claims, and to pass this case to issue.

Status of the Claims

Claims 1-9 and 11-12 are under consideration in this application. Please cancel claims 10, 13-15 without prejudice or disclaimer. Claims 1, 3, 5-6 and 12 are being amended, as set forth in the above marked-up presentation of the claim amendments, in order to more particularly define and distinctly claim applicants' invention.

Additional Amendments

The claims are being amended to correct formal errors and/or to better recite or describe the features of the present invention as claimed. All the amendments to the claims are supported by the specification. Applicants hereby submit that no new matter is being introduced into the application through the submission of this response.

Formality Rejections

The Examiner rejected claims 1-15 under 35 U.S.C. § 112, first paragraph, for reciting claims having subject matter that is not described in the specification in a manner that will enable a skilled person in the art to make or use the invention. This rejection is outlined on pp. 2 - 4 of the Office Action. Also, the Examiner rejected claims 1-15 under 35 U.S.C. § 112, second paragraph, as being vague and indefinite. Details of these rejections are noted on pages 5 - 6 of the Office Action. Further, the Examiner objected to claim 10 and to the specification for various formal errors.

The method for assembling nucleic acid base sequences of the invention, as now recited in claim 1 (e.g., Fig. 5), comprises the steps of: providing a plurality of nucleic acid base sequences; moving a window 105 of a fixed length (e.g., of 10-32 nucleic acid bases) along a first nucleic acid base sequence 104 of the plurality of nucleic acid base sequences to

define a first fixed-length partial sequence 501 and simultaneously searching for a second nucleic acid base sequence 502 among the plurality of nucleic acid base sequences which has a second fixed-length partial sequence at a terminal region thereof exactly matching (e.g., p. 30, line 18) with the first fixed-length partial sequence 501 defined by the window 105; determining whether the second nucleic acid base sequence 502 searched in said moving step and the first nucleic acid base sequence 104 can be assembled or not by comparing a sequence (e.g., ~ 503 minus 501 on 104 in Fig. 5) adjacent to said first fixed-length partial sequence 501 of said first nucleic acid base sequence 104 with a sequence (e.g., ~ 503 minus 501 on 502) adjacent to said second fixed-length partial sequence of the second nucleic acid base sequence 502 to be sufficiently similar via a high speed algorithm (p. 16, lines 7-8); and assembling said first nucleic acid base sequence and said second nucleic acid bases sequence if the second nucleic acid base sequence and the first nucleic acid base sequence are determined to be assembled.

The invention is also directed to a method recited in claim 3 which further introduces a table (Fig. 4; p. 14, lines 12-15) by entering identification information of each of the plurality of nucleic acid base sequences and a respective fixed-length partial sequence located in a terminal region of each of the nucleic acid base sequences thereinto.

The invention is also directed to a method recited in claim 5 (Fig. 2; p.p. 13-16) which further introduces a step of sorting a plurality of nucleic acid base sequences in descending order of sequence lengths, a step of selecting one of the nucleic acid base sequences with longest sequence length as the first consensus sequence, repeating the fourth step to the sixth step are until said fixed length window completes the scanning throughout said first consensus sequence, and repeating said third step to said sixth step until all of the plurality of nucleic acid base sequences are selected in the fourth step and compared in the fifth step.

Claim 9 recites a step of specifying an upper limit c as an expected number of entries retrieved from said table of an identical fixed-length partial sequence located in different nucleic acid base sequences or different positions in the nucleic acid base to be assembled to said first consensus sequences ("*The user can input and specify an upper limit c of an expected value of the number of entries which are found coincidentally despite lack of the true overlap at the time of referring the fixed-length partial sequence table 103 into the inputting and displaying area 1008 in the part 1022 for setting the fixed-length partial sequence length.*" p. 28, lines 1-6).

Regarding the alleged new matter "comparing a sequence adjacent to said first fixed-length partial sequence of said first nucleic acid base sequence with a sequence adjacent to said second fixed-length partial sequence of the second nucleic acid base sequence to be sufficiently similar via a greedy alignment algorithm," Applicants respectfully contend that it fully support by the following direct citation from the specification as "a greedy alignment algorithm" is being amended into "a high speed algorithm". "*When it has been found that a partial sequence 106 of a certain input sequence completely matches with a sequence defined by a fixed length window 105 as a result of referring to the table, whether it is included or not in the same cluster is verified by the detailed comparison of the sequences at the overlapping portion. Then members are included in the cluster one after another, based on a* ***greedy method*** (p. 12, lines 20-26)." "*In this sequence comparison, a position of the exact matching whose length is between the consensus sequence and the input sequence is apparent, so that* ***a high speed algorithm described in Zhang****, Z. et al., J. Comput. Biol., 7 (1-2): 203-14, 2000 is used* (p. 16, lines 5-9)." On page 206, third paragraph of Zhang's article (submitted via IDS), "*Greedy alignment algorithms work directly with a measurement of the difference between two sequences, rather than their similarity. In other words, near-identity of sequences is characterized by a small positive number instead of a large one. In the simplest approach, an alignment is assessed by counting the number of its differences, i.e., the number of columns that do not align identical nucleotides. The distance, D(i, j), between the strings a1a2 . . . ai and b1b2 . . . b j is then defined as the minimum number of differences in any alignments of those strings.*" "*The clustering and assembling are performed by repeatedly processing this procedure based on* ***greedy method*** *until no unprocessed input nucleic acid base sequence is left* (Abstract)."

Regarding the outstanding ennoblement rejection against the determining step, Applicants contend the recitation in independent claims 1, 3, 5 in conjunction with Zhang's article allows one skilled in the art to "determining whether the second nucleic acid base sequence 502 searched in said moving step and the first nucleic acid base sequence 104 can be assembled or not." In particular, how to move a window of a fixed-length to search for a second nucleic acid base sequence is described in line 16-24 of page 15 "*the fixed length window 105 having a width s is allowed to move through the whole consensus sequence 104 of the cluster. While moving the window, the fixed-length partial sequence table 103 is referred to by using the partial sequence defined by the window as a key, and a candidate for the input sequence which becomes a potential member of the cluster is searched (Step 204 in FIG. 2).*"

How to choose the length of a "fixed-length partial sequence" is described in line 9 of page 13 to line 8 of page 14. "*Next, the process proceeds to Step 202 in FIG. 2 and constructs a fixed-length partial sequence table 103. When constructing the fixed-length partial sequence table 103, partial sequences 102 having a length of s at the head and tail ends of all the input sequences 101 is entered into the table 103 as shown in FIG. 4. If the length s of the partial sequence is taken longer, the probability of occurrence of coincidence between the lengths s can be decreased regardless of the presence of a true overlap between the input sequences, so that the processing time can be shorten. However, if the length s of the partial sequence is excessively taken too long, the sensitivity for searching for an overlap will become lower. In the present invention, the value s has a lower limit which is represented by an expression (1) described below, in order to shorten the processing time.*

$$s \geq \frac{1}{2}\log\frac{KN}{c} \qquad \cdots(1)$$

In the above expression (1), N is the number of input sequences, K is the number of partial sequences selected from each sequence, and c is a parameter given by a user and is an amount specifying an upper limit of the expected value of the number of exact matching which can be found after each reference to the fixed-length partial sequence table 103 regardless of the presence of the true overlap between the input sequences. If the value c becomes larger, the value s can be smaller. Thus the length of the partial sequence becomes shorter, so that the sensitivity for searching for an overlap can be higher. However, the computing time for processing the coincidence matching becomes longer, so that the processing speed decreases. In this specification, the base of logarithms is 2." Further more, *s* should be no less than 10 to deal with a dataset that simulate practical applications (p. 35, line 6-12) and no more than 16 or 32 (p. 18, line 5-8).

How to determine whether the second nucleic acid base sequence can be assembled is described in line 25 of page 15 to line 9 of page 16 "*Suppose that an exact matching 501 with a certain input sequence 502 is found when referring to the fixed-length partial sequence table 103. Only the occurrence of the exact matching 501 having a length of s is not sufficient as a condition for adding this sequence 502 to the cluster because this exact matching may occur merely by coincidence. Therefore, it should be verified that both of the entire overlapping portions 503 are sufficiently similar to each other and the assembling is possible without contradiction between them by comparing one sequence with the other (Step 205 in FIG. 2). In this sequence comparison, a position of the exact matching whose length is s between the*

consensus sequence and the input sequence is apparent, so that a high speed algorithm described in Zhang, Z. et al., J. Comput. Biol., 7 (1-2): 203-14, 2000 is used." For example, counting a number of different nucleic acid bases thereof (p. 206, 3rd and 4th paragraphs of Zhang)., then assembling said first nucleic acid base sequence and said second nucleic acid bases sequence if the number of different nucleic acid bases of the second nucleic acid base sequence and the first nucleic acid base sequence is smaller than a value determined by an user (p. 206, 4th paragraph).

How to assemble the first and the second nucleic acid sequences is described in line 10-19 of page 16 "*If it is determined by the sequence comparison of Step 205 that both sequences within the entire overlapping portions 503 are well similar to each other, the input sequence 502 is added to the cluster and the consensus sequence 104 is also modified into a new consensus sequence 504 (Step 206 in FIG. 2). An extended portion 505 of the consensus sequence is also included within a moving area of the fixed length window 105 having a width of s. An entry in the fixed-length partial sequence table 103, which is associated with the input sequence 502 being added to the cluster, is deleted.*"

Accordingly, the withdrawal of the outstanding ennoblement rejection is in order, and is therefore respectfully solicited.

Regarding the limitation of "fixed-length partial sequence," it is recited as 10-32 nucleic acid base long in claim 8 as an example. Applicants contend that there in no need to specify the length since it varies depending on demands of a user, i.e., one skilled in the art .

Regarding the limitation of "sufficiently similar," it is related to how to use the "high speed algorithm (p. 16, line 8-9). The algorithm determines bases of two sequences that correspond. To determine whether two sequences are "sufficiently similar" is to choose a threshold on similarity. Applicants contend that there in no need to specify the threshold on similarity since it varies depending on demands of a user, i.e., one skilled in the art who is familiar with this kind of sequence comparison techniques.

Regarding the rejection of missing essential steps, recitation of "counting a number of different nucleic acid bases thereof via a high speed algorithm" and "if the number of different nucleic acid bases of the second nucleic acid base sequence and the first nucleic acid base sequence is smaller than a value determined by an user" are being added to claims 1, 3, 5.

Regarding the word "proceed" in claim 10, the rejection becomes moot as the calm is being cancelled without prejudice or disclaimer.

In addition, the relevant support in the specification for claims 2-12 are listed as follows:

Claim 2: the first nucleic acid base sequence replicated into the consensus sequence is described in line 12-14 of page 15, the consensus sequence modified into a new consensus sequence 504 is described in line 12-14 of page 16, and the new consensus sequence to be used during the repetition is described in line 20-21 of page 16.

Claim 3: how to enter identification information and fixed-length partial sequences into a table is described in line 9-24 of page 14, how to construct the first consensus sequence is described in line 12-14 of page 15, how to search for a second nucleic acid base sequence is described in line 16-24 of page 15, how to compare the first and the second nucleic acid sequences is described in line 25 of page 15 to line 9 of page 16, how to determine whether the second nucleic acid base sequence can be assembled is described in line 25 of page 15 to line 9 of page 16, and how to assemble the first and the second nucleic acid sequences so as to reconstruct the first consensus sequence is described in line 10-19 of page 16. The way to assemble them is trivial with the result of the high speed algorithm referred in line 8-9 of page 16.

Claim 4: how to select a sequence whose base length is the longest is described in line 8-13 of page 15. All provided sequences are sorted in descending order of their sequence lengths in advance (line 1-2 of page 13).

Claim 5: the first step is supported by the description in line 1-8 of page 13, the second step is supported by the description in line 9-24 of page 14, wherein the length of fixed-length partial sequences are chosen as described in line 9 of page 13 to line 8 of page 14, and further we mentioned it should be no less than 10 to deal with a dataset that simulate practical applications (line 6-12 of page 35) and no more than 16 or 32 (line 5-8 of page 18), the third. step is supported by the description in line 6-15 of page 15, the fourth step is supported by the description in line 16-24 of page 15, the fifth step is supported by the description in line 25 of page 15 to line 9 of page 16, wherein "greedy alignment algorithm" actually means the "high speed algorithm" referred in line 8 of page 16, the sixth step is supported by the description in line 10-1 .9 of page 16; and the repetition recited in the final paragraph of claim 5 is supported by the description in line 20-25 of page 16.

Claim 6: the step of picking up more than two fixed-length partial sequences is described in line 11-26 of page 18.

Claim 7: the step of designating a range of the terminal region is described in line 7-

12 of page 27.

Claim 8: the value of. s is limited to be no less than 10 and no more than 32. The lower bound is supported by the description in page 34-35, in particular in line 6-12 of page 35. 'the size of dataset for the experiment described in the specification is chosen to simulate the clustering and assembling; of SSTs derived front human mRNA sequences as described in line 7-15 of page 34. The upper bound is chosen so that each of said fixed-length partial sequences can be encoded in two computing words, which is supported by the description in line 5-8 of page 18.

Claim 9: the step of specifying a length s as an integer satisfying the expression (1) is supported by the description in line 9 of page 13 to line 8 of, page 14.

Claim 10: the phrase "two-way list" is the name of a data structure used for speeding up, and is not intended to indicate two ways of 5'- and 3'-. The Two-way list is also known as doubly linked list. In the method for assembling nucleic acid sequences, as recited in claim 10, the use of two-way lists is supported by the description in line 1-9 of page 17.

Claim 11: conversion of fixed-length partial sequences into a fixed number of computing; words is supported by the description in line 1-10 of page 18.

Claim 12: limitation on the frequency of the fixed-length partial sequences to be removed from the said table is supported by the description in line 25 of page 14 to line 5 of page 15.

Accordingly, the withdrawal of the outstanding informality rejections is in order, and is therefore respectfully solicited.

Conclusion

Favorable reconsideration of this application is respectfully solicited. Should there be any outstanding issues requiring discussion that would further the prosecution and allowance of the above-captioned application, the Examiner is invited to contact the Applicant's undersigned representative at the address and phone number indicated below.

Respectfully submitted,

Stanley P. Fisher
Registration Number 24,344

Juan Carlos A. Marquez
Registration Number 34,072

REED SMITH LLP
3110 Fairview Park Drive, Suite 1400
Falls Church, Virginia 22042
(703) 641-4200

December 7, 2004

SPF/JCM/JT